



Yes, we can! Mining Arguments in 50 Years of US Presidential Campaign Debates

Shohreh Haddadan, Elena Cabrio, Serena Villata

► To cite this version:

Shohreh Haddadan, Elena Cabrio, Serena Villata. Yes, we can! Mining Arguments in 50 Years of US Presidential Campaign Debates. ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Jul 2019, Florence, Italy. pp.4684-4690. hal-02194887

HAL Id: hal-02194887

<https://inria.hal.science/hal-02194887>

Submitted on 21 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Yes, we can!

Mining Arguments in 50 Years of US Presidential Campaign Debates

Shohreh Haddadan
University of Luxembourg
Luxembourg

shohreh.haddadan@uni.lu

Elena Cabrio and Serena Villata
Université Côte d’Azur, CNRS, Inria, I3S
France

firstname.surname@unice.fr

Abstract

Political debates offer a rare opportunity for citizens to compare the candidates’ positions on the most controversial topics of the campaign. Thus they represent a natural application scenario for Argument Mining. As existing research lacks solid empirical investigation of the typology of argument components in political debates, we fill this gap by proposing an Argument Mining approach to political debates. We address this task in an empirical manner by annotating 39 political debates from the last 50 years of US presidential campaigns, creating a new corpus of 29k argument components, labeled as *premises* and *claims*. We then propose two tasks: (1) identifying the argumentative components in such debates, and (2) classifying them as *premises* and *claims*. We show that feature-rich SVM learners and Neural Network architectures outperform standard baselines in Argument Mining over such complex data. We release the new corpus USElecDeb60To16 and the accompanying software under free licenses to the research community.

1 Introduction

Political debates are public interviews where the candidates of elections are requested to confront each other on topics such as unemployment, taxes, and foreign policy. During presidential elections in the US, it is customary for the main candidates of the two largest parties, i.e., the Democratic and the Republican Parties, to engage in a debate around the most controversial issues of the time. Such debates are considered as a *de facto* election process, and in some cases they have nearly decided the outcomes of the elections (Coleman et al., 2015).

Given the importance of these debates and their innate argumentative features, they represent

a natural playground for Argument(ation) Mining (AM) methods (Peldszus and Stede, 2013; Lippi and Torroni, 2016b; Cabrio and Villata, 2018). AM deals with analyzing argumentation in various domains, such as legal cases (Mochales and Moens, 2011), persuasive essays (Stab and Gurevych, 2017), clinical trials (Mayer et al., 2018) and scientific articles (Teufel et al., 2009). The ability of identifying argumentative components and predicting their relations in such a kind of texts opens the door to cutting-edge tasks like fallacy detection, fact-checking, and counter-argumentation generation.

Despite the plethora of existing approaches and annotated corpora for AM, very few of them tackle the issue of mining argumentative structures from political debates (Lippi and Torroni, 2016a; Menini et al., 2018; Duthie and Budzynska, 2018; Visser et al., 2019). To the best of our knowledge, none of them take on the identification of argument components (i.e., *premises* and *claims*) on a large corpus of political debates. This paper fills this gap by (1) performing a large-scale annotation study over 50 years of US presidential campaigns from 1960 (Nixon vs. Kennedy) to 2016 (Trump vs. Clinton), resulting in 29k annotated argument components, and (2) experimenting with feature-rich SVM learners and neural architectures outperforming standard baselines in Argument Mining. Finally, to ensure full reproducibility of our experiments, we provide all data and source codes under free licenses.

The paper is organized as follows. Section 2 discusses the related work and compares it to the proposed approach. In Section 3, we present the corpus of political debates we built along with some examples from the annotation guidelines. Section 4 describes the experimental setting, reports on the obtained results and discusses the main errors which occurred.

2 Background and related work

AM is “the general task of analyzing discourse on the pragmatics level and applying a certain argumentation theory to model and automatically analyze the data at hand” (Habernal and Gurevych, 2017). Two tasks are crucial in Argument Mining: *i) Argument component detection* in the input text: this step may be further split in the detection of argument components (i.e., claims and premises) and of their textual boundaries. Different methods have been tested, like Support Vector Machines (SVM) (e.g., (Lippi and Torroni, 2016c)), Naïve Bayes classifiers (Duthie et al., 2016), Logistic Regression (Levy et al., 2014) and Neural Networks (Stab and Gurevych, 2017), and *ii) Prediction of the relations* holding between the argumentative components (i.e., *attacks* and *supports*). Relations can be predicted between arguments (Cabrio and Villata, 2013) or between argument components (Stab and Gurevych, 2017).

Regarding political debates, (Menini et al., 2018) predict relations on speeches of the Nixon-Kennedy campaign considering only annotations on the relations among such arguments. (Lippi and Torroni, 2016a) focus on the 2015 UK election debates to study the impact of vocal features from the speeches on the identification of claims in debates. They built a small corpus of political debates annotated with premises and claims. (Duthie and Budzynska, 2018) proposed the *ethos mining* task aiming at detecting ethotic arguments and the relations among the politicians and the parties in the UK Parliament. Sentences are annotated as being ethotic arguments or not. (Basave and He, 2016) studied the use of semantic frames for modelling argumentation in speakers’ discourse. They investigated the impact of argumentation as a influence rank indicator for politicians on the 20 debates for the Republican primary election. Finally, (Visser et al., 2019) present a dataset composed of the transcripts of televised political debates leading up to the 2016 presidential election in the US, with the addition of the reactions from the social media platform Reddit. The corpus is annotated based on the Inference Anchoring Theory, and not with argument components. Contrary to past works, we create a huge annotated dataset including 39 political debates, and we present a successful attempt to argument component detection on such a big corpus of political debates.

Year	Candidates	T	S	W
1960	Kennedy-Nixon	255	2082	48326
1976	Carter-Ford	270	1874	46444
1980	Anderson-Reagan	200	1141	28765
1984	Mondale-Reagan	365	2376	50126
1988	Bush-Dukakis	484	2599	52780
1992	Bush-Clinton-Perot	929	4057	73688
1996	Clinton-Dole	280	2299	32088
2000	Bush-Gore	564	3225	71852
2000	Cheney-Lieberman	106	835	16395
2004	Bush-Kerry	419	3487	55486
2004	Cheney-Edwards	169	1069	20486
2008	McCain-Obama	505	2829	56379
2012	Obama-Romney	676	2352	62097
2012	Biden-Ryan	425	1252	20785
2016	Clinton-Trump	954	2536	40530
TOT.		6601	34013	676227

Table 1: Statistics on the debate transcripts: number of speech turns (T), sentences (S) and words (W).

3 Dataset creation

The *USElecDeb60To16 v.01* dataset was collected from the website of the Commission on Presidential Debates¹, which provided transcripts of the debates broadcasted on TV and held among the leading candidates for the presidential and vice presidential nominations in the US. *USElecDeb60To16* includes the debates starting from Kennedy and Nixon in 1960 until those between Clinton and Trump in 2016. Table 1 provides some statistics on the dataset in terms of number of turns in the conversations, of sentences and of words in the transcripts. The unique properties of this dataset are its size (see Table 1), its peculiar nature of containing reciprocal discussions (mainly between Democrats and Republicans), and its time line structure. The motivation for creating a new corpus is twofold: *i)* to the best of our knowledge, no other big corpus on political debates annotated at a argument component level for Argument Mining exists, and *ii)* we ensure the reproducibility of the annotation, writing guidelines, inspired from (Rinott et al., 2015; Lippi and Torroni, 2016a), with precise rules for identifying and segmenting argument components (i.e., claims and premises) in political debates.²

In the following, we detail the annotation of the argument components through examples from the *USElecDeb60To16* dataset.

¹<http://www.debates.org>

²The *USElecDeb60To16 v.01* dataset and the annotation guidelines are available here: <https://github.com/ElecDeb60To16/Dataset>.

Claims. Being them the ultimate goal of an argument, in the context of political debates, claims can be a policy advocated by a party or a candidate to be undertaken which needs to be justified in order to be accepted by the audience. In Example 1,³ Bush is defending the decisions taken by his administration by claiming that his policy has been effective. Claims might also provide judgments about the other candidate or parties (Example 2).

1. **Bush-Kerry, September 30, 2004:**

BUSH: My administration started what's called the Proliferation Security Initiative. Over 60 nations involved with disrupting the trans-shipment of information and/or weapons of mass destruction materials. And **[we've been effective]**. *[We busted the A.Q. Khan network. This was a proliferator out of Pakistan that was selling secrets to places like North Korea and Libya]. [We convinced Libya to disarm].*

2. **Kennedy-Nixon, September 26, 1960:**

NIXON: **[I believe the programs that Senator Kennedy advocates will have a tendency to stifle those creative energies], [I believe in other words, that his program would lead to the stagnation of the motive power that we need in this country to get progress].**

3. **Kennedy-Nixon, October 13, 1960:**

NIXON: Senator Kennedy's position and mine completely different on this. **[I favor the present depletion allowance]**. *[I favor it not because I want to make a lot of oil men rich], but because [I want to make America rich]. Why do we have a depletion allowance? Because [this is the stimulation, the incentive for companies to go out and explore for oil, to develop it].*

Taking a stance towards a controversial subject, or an opinion towards a specific issue is also considered as a claim (e.g., "I've opposed the death penalty during all of my life"). The presence of discourse indicators (e.g., "in my opinion", "I believe") is generally a useful hint in finding claims that state opinions and judgments.

Premises. Premises are assertions made by the debaters for supporting their claims (i.e., reasons or justifications). A type of premise commonly used by candidates is referring to past experience: more experienced candidates exploit this technique to assert that their claims are more relevant than their opponents because of their past experience (Example 4).

4. **Carter-Ford, September 23, 1976:**

CARTER: *[Well among my other experiences in the*

past, I've - I've been a nuclear engineer, and did graduate work in this field]. [I think I know the - the uh capabilities and limitations of atomic power].

Statistics are very commonly used as evidence to justify the claims (Example 6). Moreover, premises may be asserted in the form of examples (in such cases, they may contain discourse indicators to introduce examples and justifications, such as "because").

5. **Nixon-Kennedy, September 26, 1960:**

NIXON: We often hear gross national product discussed and in that respect may I say that *[when we compare the growth in this administration with that of the previous administration that then there was a total growth of eleven percent over seven years]; [in this administration there has been a total growth of nineteen percent over seven years]. [That shows that there's been more growth in this Administration than in its predecessor].*

6. **Clinton-Dole, October 6, 1996:**

CLINTON: *[We have ten and a half million more jobs, a faster job growth rate than under any Republican administration since the 1920s]. [Wages are going up for the first time in a decade]. [We have record numbers of new small businesses]. [We have the biggest drop in the number of people in poverty in 27 years]. [All groups of people are growing]. [We had the biggest drop in income inequality in 27 years in 1995]. [The average family's income has gone up over \$1600 just since our economic plan passed]. So [I think it's clear that we're better off than we were four years ago].*

Three expert annotators defined the annotation guidelines, then three other annotators carried out the annotation task relying on such guidelines. Each transcript has been independently annotated by at least two annotators⁴. 86% of the sentences, which were annotated at least with one component, were tagged with only one argument component, while the remaining 14% with more than one component (7% with both claims and premises).⁵ Only 0.6% of the dataset contains cross-sentence annotations (i.e., annotations which are not bound in one sentence). 19 debates have been independently annotated by three annotators to measure the IAA. The observed agreement percentage and IAA at sentence-level (following (Stab and Gurevych, 2014)) are respectively 0.83% and $\kappa = 0.57$ (moderate agreement) for argumentative-non argumentative sentences, and 63% and $\kappa = 0.4$ (fair agreement) for the argument components. Such annotation tasks are very difficult with political debates. In many examples, the choice between a premise and a claim is hard to define. In

⁴We used the Brat annotation tool (Stenetorp et al., 2012).

⁵A component cannot be both a claim and a premise (see Guidelines).

³In the examples, claims are marked in **bold**, premises in *Italics* and the component boundaries by [square brackets].

Example 7, the sentence “the way Senator [...]” is used as a premise for the previous claim, but if observed out of this context, it can be identified as a claim. This justifies the IAA on the argument component annotation.

7. **McCain-Obama, October 15, 2008:**

OBAMA: [I disagree with Senator McCain in how to do it], because [the way Senator McCain has designed his plan, it could be a giveaway to banks if we’re buying full price for mortgages that now are worth a lot less]?

To release a consistent dataset, in the reconciliation phase we computed the IAA of the annotators with two other expert annotators with background in computational linguistics on a sample of 6 debates. In case of disagreement among the first three annotators, the annotation provided by the annotator which showed to be consistently in line with the expert annotators (i.e., with a higher IAA) was included in the released dataset.

After the reconciliation phase, the *USElecDeb60To16* dataset contains the annotation of 29521 argument components (i.e., 16087 claims and 13434 premises). Notice that the number of claims is higher than the number of premises, because in political speeches the candidates make arguments mostly without providing premises for their claims. Moreover, the candidates use longer sentences (more words) to express their premises than their claims.

For our experiments, we split the dataset into train (13894 components), validation (6577 components) and test (9050 components) sets, keeping the same component distribution as in the original dataset.

4 Experimental setting

We address the argument component detection task as two subsequent classification steps, i.e., the argumentative sentences detection (Task 1), and the argumentative components identification (Task 2). We address both of these classification tasks at the sentence level (e.g., we label a sentence according to the longest component annotated in the sentence).

Methods For Task 1, we trained both a linear-kernel SVM with stochastic gradient descent learning method using bag of words features only, and a SVM classifier with rbf kernel (python scikit-learn v0.20.1, penalty parameter=10) using the features listed below to distinguish argu-

mentative sentences (i.e., sentences which contain at least one argument component) from the non-argumentative ones. For comparison, we also tested a Neural Network structured with two bidirectional LSTM layers (Hochreiter and Schmidhuber, 1997) using word embeddings from FastText (Joulin et al., 2016; Mikolov et al., 2018) as the weights for the embedding layer. The output layer determines the class Argumentative/Non-Argumentative for the input sentence. A feed-forward Neural Network was also trained using the same sentence-based features used with the SVM classifier. This network consists of two hidden layers with 64 and 32 neurons for the 1st and 2nd hidden layer, respectively.

As for the component classification step, we applied the same classifiers as for Task 1 (SVM and LSTM). For both tasks, we implemented the majority baseline for argument component classification used in (Stab and Gurevych, 2017).

We considered the following features: tf-idf of each word, NGram (bigrams and trigrams), POS of adverbs, adjectives (used by debaters to stress the correctness of their premises), different tenses of verbs and modal verbs (they often affect the certainty of the assertions, hence would be a hint of facts/non-facts in discerning between argument components), syntactic features (constituency parse trees, dept of the parsing tree), discourse connectives (and their position), NER (debaters often mention party members, former presidents, organizations and dates or numbers like statistics as examples to strengthen the premises for the claims), semantic features (sentiment polarity of the argument component and of its covering sentence (Menini et al., 2018)).

Evaluation Tables 2 and 3 present the results obtained on detecting argumentative sentences (Task 1) and classifying argumentative components (Task 2), respectively. Results obtained with linear-kernel SVM significantly outperformed the majority baseline in both tasks. Enriching the feature-set increased the classification performances by 9% on Task 1 using the rbf-kernel SVM, while only by 2.2% on Task 2. Running ablation tests for features analysis, we noticed that the lexical features (tf-idf and NGram features) strongly contribute to performance increase in both tasks. NER features – selected on the assumption that they would have improved the detection of premises as candidate tend to use

NERs to provide examples – showed to be more effective in Task 1 only. Sentiment and discourse indicator features did not show to be effective in either classification tasks. Results obtained by LSTM with word-embedding as features in both tasks are comparable to that of the SVM using all the features, showing the efficiency of neural classifiers on AM tasks using less dimensionality for the input data.

Classifier	Class	Precision	Recall	F-Score
Majority baseline	Arg	0.681	1.000	0.810
	None	0.000	0.000	0.000
	Average	0.463	0.681	0.551
SVM Linear Kernel BOW	Arg	0.758	0.980	0.855
	None	0.886	0.335	0.486
	Average	0.799	0.774	0.737
SVM Rbf Kernel All features	Arg	0.855	0.986	0.916
	None	0.834	0.293	0.433
	Average	0.851	0.853	0.823
LSTM network word-embeddings features	Arg	0.882	0.946	0.913
	None	0.668	0.463	0.547
	Average	0.841	0.854	0.843
Feed Forward Network All features	Arg	0.885	0.859	0.872
	None	0.471	0.528	0.498
	Average	0.805	0.796	0.800

Table 2: Classification results on Task 1.

Classifier	Class	Precision	Recall	F-Score
Majority baseline	Claim	0.51	1.00	0.68
	Premise	0.00	0.00	0.00
	Average	0.26	0.51	0.35
SVM Linear Kernel BOW	Claim	0.625	0.757	0.685
	Premise	0.682	0.534	0.599
	Average	0.653	0.647	0.643
SVM Rbf Kernel All features	Claim	0.631	0.830	0.717
	Premise	0.728	0.484	0.581
	Average	0.678	0.662	0.651
LSTM network word-embeddings	Claim	0.848	0.810	0.829
	Premise	0.683	0.739	0.710
	Average	0.673	0.673	0.673
Feed Forward Network All features	Claim	0.639	0.697	0.667
	Premise	0.644	0.581	0.611
	Average	0.641	0.641	0.640

Table 3: Classification results on Task 2.

Given the complexity of the task, we computed the human upper bound as the average F-score of the annotation agreement between annotators and the gold-standard. It resulted in 0.87 and 0.75 on argumentative vs. non-argumentative, and 0.74 and 0.65 on claims and premises, respectively.

Error Analysis Argumentative sentences are rarely misclassified, which results in high recall on argumentative sentence identification. Some patterns can be identified from the misclassified non-argumentative sentences. One of these patterns appears in very short non-argumentative sentences which contain an argument indicator such

as “so”, for instance the sentence: “So what should we do?” is classified as argumentative although in the context it is considered a non-argumentative sentence. Since indicators for claims are more numerous in these debates, this misclassification mostly occurs when a claim-indicator is uttered by the candidate in a non-argumentative manner.

In other cases, candidates make final remarks phrasing their speech with a structure similar to argumentative sentences, for example: “I think when you make that decision, it might be well if you would ask yourself, are you better off than you were four years ago?”.

Misclassification between claims and premises, instead, is primarily due to the fact that the component classification is highly dependent on the structure of the argument.

5 Conclusion

We investigated the detection of argument components in the US presidential campaign debates: *i)* providing a manually annotated resource of 29k argument components, and *ii)* evaluating feature-rich SVM learners and Neural Networks on such data (achieving $\sim 90\%$ w.r.t. human performance). We highlighted the strengths (e.g., satisfactory performances on different oratory styles across time and topics) and weaknesses (e.g., no argument boundaries detection on a clause level, the context of the whole debates is not considered).

For future work, we plan to *i)* automatically predict relations between argument components in the *USElecDeb60To16* dataset, and *ii)* propose a new task, i.e., *fallacy detection* so that common fallacies in political argumentation (Zurloni and Anolli, 2010) can be automatically identified, in line with the work of (Habernal et al., 2018).

Acknowledgments

Shohreh Haddadan acknowledges the support of the Luxembourg National Research Fund (FNR) (10929115). Elena Cabrio and Serena Villata have been partially supported by EU H2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 690974 for the project MIREL: MIning and REasoning with Legal texts.

The experiments presented in this paper were carried out using the HPC facilities of the University of Luxembourg (Varrette et al., 2014) – see <https://hpc.uni.lu>.

References

- Amparo Elizabeth Cano Basave and Yulan He. 2016. [A study of the impact of persuasive argumentation in political debates](#). In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 1405–1413. The Association for Computational Linguistics.
- Elena Cabrio and Serena Villata. 2013. [A natural language bipolar argumentation approach to support users in online debate interactions](#). *Argument & Computation*, 4(3):209–230.
- Elena Cabrio and Serena Villata. 2018. Five years of argument mining: a data-driven analysis. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 5427–5433.
- Stephen Coleman, J Blumler, G Moss, and M Homer. 2015. The 2015 televised election debates; democracy on demand? *Leeds University Press*.
- Rory Duthie and Katarzyna Budzynska. 2018. A deep modular RNN approach for ethos mining. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 4041–4047.
- Rory Duthie, Katarzyna Budzynska, and Chris Reed. 2016. [Mining ethos in political debate](#). In *Computational Models of Argument - Proceedings of COMMA 2016, Potsdam, Germany, 12-16 September, 2016*, volume 287 of *Frontiers in Artificial Intelligence and Applications*, pages 299–310. IOS Press.
- Ivan Habernal and Iryna Gurevych. 2017. [Argumentation mining in user-generated web discourse](#). *Comput. Linguist.*, 43(1):125–179.
- Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. 2018. [Before name-calling: Dynamics and triggers of ad hominem fallacies in web argumentation](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 386–396. Association for Computational Linguistics.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, H  rve J  gou, and Tomas Mikolov. 2016. Fasttext.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*.
- Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pages 1489–1500. ACL.
- Marco Lippi and Paolo Torroni. 2016a. Argument mining from speech: Detecting claims in political debates. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 2979–2985.
- Marco Lippi and Paolo Torroni. 2016b. [Argumentation mining: State of the art and emerging trends](#). *ACM Trans. Internet Techn.*, 16(2):10.
- Marco Lippi and Paolo Torroni. 2016c. Margot: A web server for argumentation mining. *Expert Systems with Applications*, 65:292–303.
- Tobias Mayer, Elena Cabrio, Marco Lippi, Paolo Torroni, and Serena Villata. 2018. Argument mining on clinical trials. *Computational Models of Argument - Proceedings of COMMA 2018, Warsaw, Poland, 12-14 September 2018*, 305:137.
- Stefano Menini, Elena Cabrio, Sara Tonelli, and Serena Villata. 2018. [Never retreat, never retract: Argumentation analysis for political speeches](#). In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 4889–4896. AAAI Press.
- Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin. 2018. Advances in pre-training distributed word representations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*.
- Raquel Mochales and Marie-Francine Moens. 2011. [Argumentation mining](#). *Artificial Intelligence and Law*, 19(1):1–22.
- Andreas Peldszus and Manfred Stede. 2013. [From argument diagrams to argumentation mining in texts: A survey](#). *Int. J. Cogn. Inform. Nat. Intell.*, 7(1):1–31.
- Ruty Rinott, Lena Dankin, Carlos Alzate, Mitesh M. Khapra, Ehud Aharoni, and Noam Slonim. 2015. Show me your evidence - an automatic method for context dependent evidence detection. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 440–450. The Association for Computational Linguistics.
- Christian Stab and Iryna Gurevych. 2014. [Annotating argument components and relations in persuasive essays](#). In *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pages 1501–1510. ACL.

- Christian Stab and Iryna Gurevych. 2017. [Parsing argumentation structures in persuasive essays](#). *Computational Linguistics*, 43(3):619–659.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. Brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107. Association for Computational Linguistics.
- Simone Teufel, Advait Siddharthan, and Colin R. Batchelor. 2009. Towards domain-independent argumentative zoning: Evidence from chemistry and computational linguistics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009, 6-7 August 2009, Singapore, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1493–1502. ACL.
- S. Varrette, P. Bouvry, H. Cartiaux, and F. Georgatos. 2014. Management of an academic hpc cluster: The ul experience. In *Proc. of the 2014 Intl. Conf. on High Performance Computing & Simulation (HPCS 2014)*, pages 959–967, Bologna, Italy. IEEE.
- Jacky Visser, Barbara Konat, Rory Duthie, Marcin Koszowy, Katarzyna Budzynska, and Chris Reed. 2019. Argumentation in the 2016 us presidential elections: annotated corpora of television debates and social media reaction. *Language Resources and Evaluation*.
- Valentino Zurloni and Luigi Anolli. 2010. [Fallacies as argumentative devices in political debates](#). In *Multi-modal Communication in Political Speech. Shaping Minds and Social Action - International Workshop, Political Speech 2010, Rome, Italy, November 10-12, 2010, Revised Selected Papers*, volume 7688 of *Lecture Notes in Computer Science*, pages 245–257. Springer.